Fall 2021

CSCI E-101 Module 4 Python Managing Data Practice Worksheet

(This is a non-graded exercise – you do not need to submit it. Solving this problem will help you to prepare for the midterm)

The Excel file "CovidTests.xlsx" contains the Covid test data of 100 Covid testing clinics collected on May 6, 2020. You want to find out if the test string for each clinic is valid. If the clinic's test string is valid, you want to find out the total number of tests performed by the clinic, the total number of positive test cases, and the total number of negative test cases.

Your Tasks:

1) Read data from CovidTests.xlsx into Python Pandas Dataframe
2) Perform the validation and calculation tasks as stated in the problem description
3) Add a column to the data frame and name it "isValid" to store the values "yes" True and "no" for False
4) Add another column to the data frame and name it "TotalTests" to store the total number of tests performed by each clinic
5) Add another column to the data frame and name it "Positive" to store the total number of positive test cases for each clinic
6) Add another column to the data frame and name it "Negative" to store the total number of negative tests cases for each clinic
7) Store the data frame with the added columns to an Excel file named "ClinicTestResults.xlsx"

NOTE:

Use the SPECS for isValidString from Python Assignment 1. Use all functions you developed for Python Assignment 1 to solve this problem. You are only allowed to use the Pandas functions we covered in class. No other import modules/libraries are allowed. You might need to fix your Python Assignment 1 code if you did not pass the 30 test cases we used to test your code.